

Contents lists available at ScienceDirect

Journal of Statistical Planning and Inference

journal homepage: www.elsevier.com/locate/jspi

Inference with the lognormal distribution

Nicholas T. Longford*

Departament d'Economia i Empresa, Universitat Pompeu Fabra, SNTL and UPF, Ramon Trias Fargas 25-27, 08005 Barcelona, Spain

ARTICLE INFO

Article history:

Received 17 June 2008

Received in revised form

21 October 2008

Accepted 24 October 2008

Keywords:

Efficiency

Lognormal distribution

Minimax estimator

Taylor expansion

ABSTRACT

Several estimators of the expectation, median and mode of the lognormal distribution are derived. They have the form $\exp(\hat{\mu} + b\hat{\sigma}^2)$, where $\hat{\mu}$ and $\hat{\sigma}^2$ are the usual estimators of the mean and variance of the underlying normal distribution, and b is a suitable constant or function of $\hat{\sigma}^2$. The estimators aim to be approximately unbiased, efficient, or to have a minimax property in the class of estimators we consider. The small-sample properties of these estimators are assessed by simulations and, when possible, analytically. Some of these estimators of the expectation are far more efficient than the maximum likelihood or the minimum-variance unbiased estimator, even for substantial sample sizes.

© 2008 Elsevier B.V. All rights reserved.

1. Introduction

The lognormal distribution is used in a wide range of applications, when the multiplicative scale is appropriate and the log-transformation removes the skew and brings about symmetry of the data distribution (Limpert et al., 2001). Normality is the preferred distributional assumption in many contexts, and logarithm is often the first transformation that an analyst considers to promote it. Linear models are convenient to specify and all the relevant moments are easy to calculate and operate with on the log scale. However, there are instances when moments, and the expectation in particular, are of interest on the original (exponential) scale. For example, the lognormal distribution is frequently applied to variables in monetary units, such as companies' assets, liabilities and profits, residential property prices (Zabel, 1999) and household income (Longford and Pittau, 2006). The population mean of such a variable may be a much more relevant target for inference than the population mean of its logarithm. The sample mean is a suitable estimator for large samples, when asymptotics provide a good approximation. In samples that are not large enough, and especially when the underlying (normal-scale) variance is large, the sample mean is very inefficient. We explore several alternatives and study their small-sample properties.

Finney (1941) derived the minimum-variance unbiased estimator of the expectation and variance of the lognormal distribution, but it involves the evaluation of an infinite series; see Thöni (1969) for an application. Aitchison and Brown (1957) is a comprehensive reference for the lognormal distribution; see also Crow and Shimizu (1988). Royston (2001) considers the lognormal distribution as an alternative basis for survival analysis, claiming robustness and convenience. He fits a linear model on the log scale, but implies that the prediction obtained on the log scale can be transformed back to the original scale straightforwardly. The confidence intervals can be, but the prediction as such cannot, because the transformation is highly nonlinear. Toma (2003) derives estimators for the multivariate lognormal distribution, but her focus is on large-sample properties. Zhou et al. (1997) study tests for comparing two lognormal samples, but consider only test statistics that resemble the t , which include the likelihood ratio. We derive a closed-form estimator that is biased but is more efficient than Finney's estimator.

The expectation, median and mode of the lognormal distribution $\mathcal{LN}(\mu, \sigma^2)$, defined by exponentiation of the normal distribution $\mathcal{N}(\mu, \sigma^2)$, are $\exp(\mu + \frac{1}{2}\sigma^2)$, $\exp(\mu)$ and $\exp(\mu - \sigma^2)$, respectively. This motivates the general problem of estimating

* Tel.: +34935422551; fax: +34935421746.

E-mail address: NTL@sntl.co.uk.

the quantity $\theta(a) = \exp(\mu + a\sigma^2)$, with a given constant a , by exploring the estimators of the form $\hat{\theta}(b_a) = \exp(\hat{\mu} + b_a\hat{\sigma}^2)$. The naive estimator $\hat{\theta}(a)$, such as $\exp(\hat{\mu} + \frac{1}{2}\hat{\sigma}^2)$ for $a = \frac{1}{2}$, is biased and inefficient, and its bias correction is ineffective. We focus on efficient estimation, but derive some results related to unbiased estimation to highlight that attempts at bias reduction and minimisation of the mean squared error (MSE) lead to very different estimators.

In the next section, we seek estimators $\hat{\theta}(b_a)$ or $\hat{\theta}(\hat{b}_a)$ that are unbiased, attain the minimum MSE and have a minimax property. The following sections describe simulations of these estimators. The paper is concluded by a discussion.

2. Estimation

Let $\mathbf{X} = (X_1, \dots, X_n)^\top$ be a random sample from $\mathcal{LN}(\mu, \sigma^2)$. The sample mean $\hat{\mu} = \{\log(X_1) + \dots + \log(X_n)\}/n$ is unbiased and efficient for μ , with sampling variance $\text{var}(\hat{\mu}) = \sigma^2/n$. However, these desirable properties are lost by nonlinear transformations; $\exp(\hat{\mu})$ is unbiased and efficient for neither $\exp(\mu)$ nor $E(X) = \exp(\mu + \frac{1}{2}\sigma^2)$. We have

$$E\{\exp(\hat{\mu})\} = \exp\left(\mu + \frac{\sigma^2}{2n}\right),$$

$$\text{var}\{\exp(\hat{\mu})\} = \exp\left(2\mu + \frac{\sigma^2}{n}\right) \left\{ \exp\left(\frac{\sigma^2}{n}\right) - 1 \right\}.$$

We assume that there is an unbiased estimator $\hat{\sigma}^2$ of σ^2 , and that

$$k \frac{\hat{\sigma}^2}{\sigma^2} \sim \chi_k^2,$$

the χ^2 distribution with k degrees of freedom, defined by the density

$$f(x) = \frac{1}{\Gamma\left(\frac{1}{2}k\right)} \left(\frac{1}{2}\right)^{(k/2)} x^{(k/2)-1} \exp\left(-\frac{x}{2}\right)$$

on $(0, +\infty)$. In the introduced setting, k is equal to $n - 1$, but we consider generalisations in which more than one degree of freedom is lost. For evaluating the bias and MSE of estimators of the form $\hat{\theta}(b) = \exp(\hat{\mu} + b\hat{\sigma}^2)$, we require expressions for the expectation and variance of $\exp(b\hat{\sigma}^2)$.

For a random variable Y with χ_k^2 distribution and scalar $c < \frac{1}{2}$,

$$E\{\exp(cY)\} = \frac{1}{\Gamma\left(\frac{1}{2}k\right)} \left(\frac{1}{2}\right)^{(k/2)} \int_0^{+\infty} x^{(k/2)-1} \exp\left\{-\frac{x}{2}(1-2c)\right\} dx$$

$$= (1-2c)^{-(k/2)}.$$
(1)

For the variance we have the identity

$$\text{var}\{\exp(cY)\} = E\{\exp(2cY)\} - [E\{\exp(cY)\}]^2$$

$$= (1-4c)^{-(k/2)} - (1-2c)^{-k},$$
(2)

which holds as long as $c < \frac{1}{4}$; otherwise the variance is not defined. Expressions for the expectation and variance of $\exp(b\hat{\sigma}^2)$ are obtained by substituting $c = b\sigma^2/k$ in (1) and (2).

2.1. Approximately unbiased estimation

We seek first the constant b for which $\hat{\theta}(b) = \exp(\hat{\mu} + b\hat{\sigma}^2)$ is unbiased for $\theta(a) = \exp(\mu + a\sigma^2)$. As $\hat{\mu}$ and $\hat{\sigma}^2$ are independent,

$$E\{\hat{\theta}(b)\} = \exp\left(\mu + \frac{\sigma^2}{2n}\right) \left(\frac{k}{k-2b\sigma^2}\right)^{k/2}.$$
(3)

Therefore $\hat{\theta}(b)$ is unbiased for $\theta(a)$ when

$$\frac{k-2b\sigma^2}{k} = \exp\left\{-\frac{2\sigma^2}{k}\left(a - \frac{1}{2n}\right)\right\},$$

that is, for

$$b_{a,ub}^* = \frac{k}{2\sigma^2} \left[1 - \exp \left\{ -\frac{2\sigma^2}{k} \left(a - \frac{1}{2n} \right) \right\} \right]. \quad (4)$$

As $b_{a,ub}^*$ depends on σ^2 , we have to estimate it. Its naive estimator is denoted by $\hat{b}_{a,ub}^*$. The dependence of $b_{a,ub}^*$ on σ^2 is avoided by the Taylor expansion, which yields the approximation

$$b_{a,ub}^\dagger = a - \frac{1}{2n}. \quad (5)$$

It can be interpreted as a multiplicative bias correction of the naive estimator $\exp(\hat{\mu})$ which has the expectation $\exp\{\mu + \sigma^2/(2n)\}$. For $a = \frac{1}{2}$ it agrees with the linear term of the expansion of the function $g(t)$ in the minimum-variance unbiased estimator $\exp(\hat{\mu})g(\frac{1}{2}\hat{\sigma}^2)$, derived by Finney (1941);

$$g(t) = 1 + \sum_{h=1}^{\infty} \frac{(n-1)^{2h-1} t^h}{h! n^h} \prod_{m=1}^{h-1} \frac{1}{n+2m-1},$$

with the convention that the product of no terms (for $h = 1$) is equal to unity. Unlike Finney's estimator, neither $\hat{\theta}(\hat{b}_{a,ub}^*)$ nor $\hat{\theta}(b_{a,ub}^\dagger)$ are unbiased because the estimators \hat{b} are not linear in $\hat{\sigma}^2$. Both estimators $\hat{\theta}$ turn out to be very inefficient for all three values of a that are of interest, $\frac{1}{2}$, 0, and -1 .

2.2. Approximately minimum MSE

The sampling variance of $\hat{\theta}(b)$ is

$$\begin{aligned} \text{var}(\hat{\mu} + b\hat{\sigma}^2) &= E\{\exp(2\hat{\mu})\}E\{\exp(2b\hat{\sigma}^2)\} - [E(\hat{\mu})E\{\exp(b\hat{\sigma}^2)\}]^2 \\ &= \exp\left(2\mu + \frac{2\sigma^2}{n}\right) \left(\frac{k}{k-4b\sigma^2}\right)^{k/2} - \exp\left(2\mu + \frac{\sigma^2}{n}\right) \left(\frac{k}{k-2b\sigma^2}\right)^k, \end{aligned} \quad (6)$$

and its bias for $\theta(a)$ is

$$\exp(\mu) \left\{ \exp\left(\frac{\sigma^2}{2n}\right) \left(\frac{k}{k-2b\sigma^2}\right)^{k/2} - \exp(a\sigma^2) \right\}.$$

Hence the MSE of $\hat{\theta}(b)$ in estimating $\theta(a)$ is

$$m(b; a) = \exp(2\mu) \left\{ \exp(2a\sigma^2) - 2 \exp\left(a\sigma^2 + \frac{\sigma^2}{2n}\right) \left(\frac{k}{k-2b\sigma^2}\right)^{k/2} + \exp\left(\frac{2\sigma^2}{n}\right) \left(\frac{k}{k-4b\sigma^2}\right)^{k/2} \right\}. \quad (7)$$

The minimum of this function of b is found as the root of its derivative

$$\frac{\partial m}{\partial b} = 2\sigma^2 \exp(2\mu) \left\{ \exp\left(\frac{2\sigma^2}{n}\right) \left(\frac{k}{k-4b\sigma^2}\right)^{(k/2)+1} - \exp\left(a\sigma^2 + \frac{\sigma^2}{2n}\right) \left(\frac{k}{k-2b\sigma^2}\right)^{(k/2)+1} \right\}.$$

The solution is

$$b_{a,ms}^* = \frac{k}{2\sigma^2} \frac{D_a - 1}{2D_a - 1}, \quad (8)$$

where

$$D_a = \exp \left\{ \frac{2\sigma^2}{k+2} \left(a - \frac{3}{2n} \right) \right\}.$$

Finiteness of the MSE ($b < \frac{1}{4}k/\sigma^2$) implies the condition $D_a > \frac{1}{2}$. Unlike in (5), the (linear) Taylor expansion of $b_{a,ms}^*$ (or D_a) does not yield a radical simplification, and the resulting estimator $\hat{\theta}(b_{a,ms}^\dagger)$ is very inefficient.

If we managed to obtain $b_{a,ms}^*$, we would attain with $\hat{\theta}(b_{a,ms}^*)$ the so-called *ideal* MSE

$$m(b_{a,ms}^*; a) = \exp(2\mu) \left\{ \exp\left(\frac{2\sigma^2}{n}\right) (2D_a - 1)^{k/2} - 2 \exp\left(\frac{\sigma^2}{2n} + a\sigma^2\right) \left(\frac{2D_a - 1}{D_a}\right)^{k/2} + \exp(2a\sigma^2) \right\}. \quad (9)$$

It is the lower bound for the MSE in the class of estimators $\hat{\theta}(b)$ in which b is a constant. We regard $m(b_{a,ms}^*; a)$ as a reference against which we compare the MSEs of other (realisable) estimators of $\theta(a)$. The estimator $\hat{\theta}(\hat{b}_{a,ms}^*)$, obtained by estimating $b_{a,ms}^*$ naively, will turn out to be the most efficient of the estimators of $\theta(a)$ that we study.

2.3. Minimax estimation

For coefficient b such that $0 < b < \frac{1}{4}k/\sigma^2$, the variance $\text{var}\{\hat{\theta}(b)\}$ is an increasing function of σ^2 . To see this, we differentiate the expression in (6). Denote by A and B the terms in (6), so that $\text{var}\{\hat{\theta}(b)\} = A - B$. Then

$$\begin{aligned} \frac{\partial \text{var}\{\hat{\theta}(b)\}}{\partial \sigma^2} &= \frac{1}{n} \exp(2\mu) \left\{ 2 \exp\left(\frac{2\sigma^2}{n}\right) \left(\frac{k}{k - 4b\sigma^2}\right)^{k/2} - \exp\left(\frac{\sigma^2}{n}\right) \left(\frac{k}{k - 2b\sigma^2}\right)^k \right\} \\ &\quad + 2b \exp(2\mu) \left\{ \exp\left(\frac{2\sigma^2}{n}\right) \left(\frac{k}{k - 4b\sigma^2}\right)^{(k/2)+1} - \exp\left(\frac{\sigma^2}{n}\right) \left(\frac{k}{k - 2b\sigma^2}\right)^{k+1} \right\} \\ &= 2 \left(\frac{1}{n} + \frac{bk}{k - 4b\sigma^2} \right) A - \left(\frac{1}{n} + \frac{2bk}{k - 2b\sigma^2} \right) B. \end{aligned}$$

For $b > 0$, the factors of A and B are both positive and their difference

$$\frac{1}{n} + \frac{2bk}{k - 4b\sigma^2} - \frac{2bk}{k - 2b\sigma^2} = \frac{1}{n} + \frac{4b^2k\sigma^2}{(k - 2b\sigma^2)(k - 4b\sigma^2)}$$

is also positive. Therefore $A > B > 0$ implies that the derivative is positive. The derivative is positive even at $\sigma^2 = 0$, where it is equal to $(2A - B)/n$. Therefore, the variance of $\hat{\theta}(b; \sigma^2 d)$ is an increasing function of σ^2 even for some b in the left-hand neighbourhood of zero. According to (8), $b_{a,ms}^* > 0$ corresponds to $a > 3/(2n)$. For the median, $b_{0,ms}^*$ is negative, but of the order $-O(1/n)$.

One can expect that the MSE $m(b, a)$ is an increasing function of σ^2 for any $b > 0$ such that $b < \frac{1}{4}k/\sigma^2$. If we cannot find an estimator that is (uniformly) efficient for all $\sigma^2 > 0$, we might pay more attention to efficiency for greater values of σ^2 , for which more is at stake. This motivates the following approach to estimating $\theta(a)$. Suppose we are certain (or very confident) that σ^2 does not exceed a specified value σ_{mx}^2 . Then we use the coefficient $b_{a,mx}$ for which the estimator $\hat{\theta}(b_{a,mx})$ is efficient when $\sigma^2 = \sigma_{mx}^2$,

$$b_{a,mx} = \frac{k}{2\sigma_{mx}^2} \frac{D_{a,mx} - 1}{2D_{a,mx} - 1}, \quad (10)$$

as in (8), with implicitly defined $D_{a,mx}$, and apply the estimator $\hat{\theta}(b_{a,mx})$. Rigour would be enhanced by writing $b_{a,mx} = b_{a,mx}(\sigma_{mx}^2)$, but that would make the notation cumbersome.

One might reasonably expect that the estimator $\hat{\theta}(b_{a,mx})$ based on a value $\sigma_{mx}^2 = \sigma_1^2$ is more efficient for all $\sigma^2 \in (0, \sigma_1^2]$ than $\hat{\theta}(b_{a,mx})$ based on a value $\sigma_2^2 > \sigma_1^2$. That is, for a sharper (smaller) upper bound σ_{mx}^2 we should be rewarded by uniformly more efficient estimation, so long as this bound is justified. This conjecture is proved by differentiating the MSE of $\hat{\theta}(b_{a,mx})$ with respect to σ_{mx}^2 .

Directly from (7) we obtain the identity

$$\begin{aligned} m(b_{a,mx}; a) &= \exp(2\mu) \left[\exp\left(\frac{2\sigma^2}{n}\right) \left\{ 1 - \frac{2(D_{a,mx} - 1)}{2D_{a,mx} - 1} \frac{\sigma^2}{\sigma_{mx}^2} \right\}^{-k/2} \right. \\ &\quad \left. - 2 \exp\left(\frac{\sigma^2}{2n} + a\sigma^2\right) \left(1 - \frac{D_{a,mx} - 1}{2D_{a,mx} - 1} \frac{\sigma^2}{\sigma_{mx}^2} \right)^{-k/2} + \exp(2a\sigma^2) \right]. \quad (11) \end{aligned}$$

Let $C_a = \log(D_{a,mx})/\sigma_{mx}^2$, so that $\partial D_{a,mx}/\partial \sigma_{mx}^2 = C_a D_{a,mx}$. For $a > 3/(2n)$, $C_a > 0$. Then

$$\begin{aligned} \frac{\partial m(b_{a,mx}; a)}{\partial \sigma_{mx}^2} &= \exp\left(2\mu + \frac{2\sigma^2}{n}\right) \frac{k\sigma^2}{\sigma_{mx}^4} \frac{\{(2D_{a,mx} - 1)(D_{a,mx} - 1) - C_a D_{a,mx} \sigma_{mx}^2\}}{(2D_{a,mx} - 1)^2} \\ &\quad \times \left[\exp\left\{\sigma^2\left(a - \frac{3}{2n}\right)\right\} \left\{ \frac{(2D_{a,mx} - 1)\sigma_{mx}^2}{(2D_{a,mx} - 1)\sigma_{mx}^2 - (D_{a,mx} - 1)\sigma^2} \right\}^{(k/2)+1} \right. \\ &\quad \left. - \left\{ \frac{(2D_{a,mx} - 1)\sigma_{mx}^2}{(2D_{a,mx} - 1)\sigma_{mx}^2 - 2(D_{a,mx} - 1)\sigma^2} \right\}^{(k/2)+1} \right]. \end{aligned} \quad (12)$$

The long fraction in the first row is positive because the inequality $D_{a,mx} - 1 > C_a \sigma_{mx}^2 > 0$, obtained from the Taylor expansion of $D_{a,mx}$ around $\sigma_{mx}^2 = 0$, implies that

$$(2D_{a,mx} - 1)(D_{a,mx} - 1) - C_a D_{a,mx} \sigma_{mx}^2 > (D_{a,mx} - 1)^2. \quad (13)$$

The exponential in the second row is equal to $D_a^{(k/2)+1}$, and so the sign of the expression in (12) is the same as the sign of

$$\begin{aligned} &D_a \{(2D_{a,mx} - 1)\sigma_{mx}^2 - 2(D_{a,mx} - 1)\sigma^2\} - (2D_{a,mx} - 1)\sigma_{mx}^2 + (D_{a,mx} - 1)\sigma^2 \\ &= (D_a - 1)(2D_{a,mx} - 1)\sigma_{mx}^2 - (2D_a - 1)(D_{a,mx} - 1)\sigma^2 \\ &= \{F(\sigma^2) - F(\sigma_{mx}^2)\}(2D_a - 1)(2D_{a,mx} - 1)\sigma^2 \sigma_{mx}^2, \end{aligned}$$

where $F(\sigma^2) = \sigma^{-2}(D_a - 1)/(2D_a - 1)$. The function F is decreasing, since

$$\begin{aligned} \frac{\partial F}{\partial \sigma^2} &= \frac{D_a C_a}{(2D_a - 1)^2} \frac{1}{\sigma^2} - \frac{D_a - 1}{2D_a - 1} \frac{1}{\sigma^4} \\ &= - \frac{(D_a - 1)(2D_a - 1) - C_a D_a \sigma^2}{(2D_a - 1)^2 \sigma^4} < 0, \end{aligned}$$

using the same argument as in (13). Therefore $F(\sigma^2) > F(\sigma_{mx}^2)$ and (12) is positive, whenever $\sigma^2 < \sigma_{mx}^2$. This concludes the proof that a smaller upper bound σ_{mx}^2 results in a uniformly more efficient estimator $\theta(b_{a,mx})$, so long as $b > 0$ and the bound is justified, that is, $\sigma^2 < \sigma_{mx}^2$.

The relative bias of an estimator is defined as the ratio of its expectation and its target,

$$B_{rel}(\hat{\theta}) = \frac{E(\hat{\theta})}{\theta(a)}.$$

For the minimax estimators with moderate or large k , $\log(B_{rel})$ can be approximated by a scalar multiple of the variance σ^2 :

$$\log\{E(\hat{\theta})\} - \log\{\theta(a)\} = \sigma^2 \left(\frac{1}{2n} - a \right) - \frac{k}{2} \log \left(1 - \frac{\sigma^2}{\sigma_{mx}^2} \frac{D_{a,mx} - 1}{2D_{a,mx} - 1} \right),$$

and, since $D_{a,mx}$ does not depend on σ^2 , the second term is well approximated by a scalar multiple of σ^2 when $D_{a,mx}$ is much greater than $\frac{1}{2}$. The approximation is not very good when estimating the mode ($a = -1$).

3. Simulations

The simulations of the various estimators of $\theta(a)$ for $a = \frac{1}{2}, 0$ and 1 described in this section are conducted for $\sigma^2 \in (0, 10)$. For orientation, a typical draw from $\mathcal{LN}(\mu, 10)$ is about $\exp(\sqrt{10}) \doteq 24$ times greater or smaller than the median $\exp(\mu)$. The densities of the distributions $\mathcal{LN}(0, \sigma^2)$ for variances $\sigma^2 = 0.1, 0.5, 2.5$ and 10 are plotted in Fig. 1. The right-hand tails of the densities are cropped. Although the density of $\mathcal{LN}(0, 10)$ attains very small values for $x > 12$, the interval $(12, +\infty)$ accounts for over 20% of the distribution. The vertical axis is also cropped; the density of $\mathcal{LN}(0, 10)$ attains its maximum of 18.72 at $x \doteq 4.5 \times 10^{-5}$.

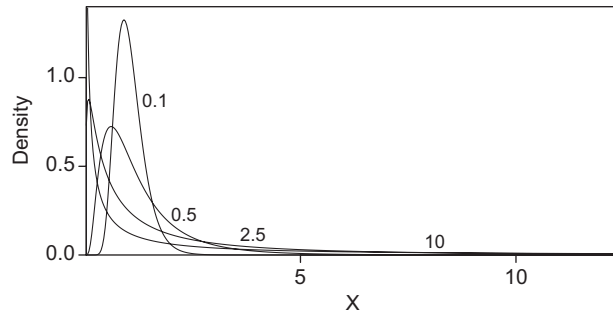


Fig. 1. The densities of the lognormal distributions based on $\mathcal{N}(0, \sigma^2)$, with σ^2 indicated in the plot.

Table 1

Estimators of the expectation of the lognormal distribution $\theta(\frac{1}{2})$.

| Criterion | Name | b_a (or \hat{b}_a) | Note ($a = \frac{1}{2}$) | |
|-------------|-------------|---|--|-----|
| Naive (CLT) | \tilde{X} | | Very inefficient for moderate σ^2 | B× |
| Naive | Naive1 | 0 | Very inefficient for small and moderate σ^2 | B× |
| Naive | Naive2 | a | Maximum likelihood; inefficient | b× |
| No bias | UnbiasedN | $\hat{b}_{a,ub}^*$; see (4) | Small bias only for small σ^2 | B× |
| No bias | UnbiasedA | $\hat{b}_{a,ub}^* = a - \frac{1}{2\pi}$ | Biased ^a | B× |
| Minimum MSE | MinMSEn | $\hat{b}_{a,ms}^*$; see (8) | Efficient | b** |
| Minimum MSE | MinMSEa | $\hat{b}_{a,ms}^*$ | Very inefficient for moderate σ^{2a} | b× |
| Finney | Finney | | Unbiased, but inefficient | U× |
| Minimax | Minmax | $b_{a,mx}$; see (10) | Used with $\sigma_{mx}^2 = 2, 4$ and 8 | b* |
| Minimum MSE | Ideal | $b_{a,ms}^*$; see (8) | Used with the value of σ^2 . The reference. | b |

The symbols at the right-hand margin indicate the following: b—bias; B—substantial bias; U—no bias; ×—inefficient estimator; *—our recommendation.

^aNot represented in Fig. 2.

The expectations and biases of all estimators have the multiplicative factor $\exp(\mu)$, and the variances and MSEs the factor $\exp(2\mu)$. Therefore we can reduce our attention to the targets $\theta(a)$ with $\mu = 0$, so that σ^2 is the sole parameter of interest. Nevertheless, μ is estimated throughout. For each estimator $\hat{\theta}$ we study the relative bias B_{rel} and relative root-MSE, defined as

$$rMSE_{rel}(\hat{\theta}) = \sqrt{\frac{MSE(\hat{\theta})}{m(b_{a,ms}^*, a)}},$$

where $m(b_{a,ms}^*, a)$ is the ideal MSE given by (9). They reduce the strong association of the bias and MSE with σ^2 .

3.1. Estimators of the mean

Table 1 lists the estimators of the expectation $\theta(\frac{1}{2}) = \exp(\mu + \frac{1}{2}\sigma^2)$. The relative biases and MSEs of these estimators with sample sizes $n = 10, 50$ and 250 (and $k = n - 1$), representing small, moderate and large samples, are plotted as functions of σ^2 in Fig. 2. They are based on simulations with 10 000 replications. The estimators $\hat{\theta}(b_{0.5,ub}^\dagger)$ and $\hat{\theta}(b_{0.5,ms}^\dagger)$ are omitted because they are uniformly less efficient than $\hat{\theta}(b_{0.5,ub}^*)$ and $\hat{\theta}(b_{0.5,ms}^*)$, respectively. The diagram shows that the estimator $\hat{\theta}(b_{a,ms}^*)$, intended to be efficient, delivers on the promise, even though it is biased throughout. For $n = 50$ and 250 , it is uniformly more efficient than even the ideal estimator $\hat{\theta}(b_{a,ms}^*)$. This is counterintuitive—not knowing the value of σ^2 is ‘rewarded’ by reduced MSE of $\hat{\theta}$. This is not a contradiction, however, because $m(b_{a,ms}^*, a)$ is the lower bound for the MSEs of estimators only in the class $\hat{\theta}(b)$, with a constant b ; $\hat{\theta}(b_{a,ms}^*)$ does not belong to this class. Finney’s estimator is unbiased, but is less efficient than the ideal, substantially so for large σ^2 . The estimators intended to be approximately unbiased, $\hat{\theta}(b_{a,ub}^*)$ and $\hat{\theta}(b_{a,ub}^\dagger)$, have substantial negative biases for all σ^2 and, for $n = 50$ and 250 , $\hat{\theta}(b_{a,ub}^*)$ approaches efficiency only for very large values of σ^2 . The pursuit of unbiasedness results in a substantial MSE inflation. For $n = 10$, $\hat{\theta}(b_{a,ub}^*)$ is efficient for large values of σ^2 , but even for $n = 15$ it is much less efficient than $\hat{\theta}(b_{a,ms}^*)$.

Despite using a smooth (analytical) reference root-MSE, some of the empirical relative root-MSEs are far from smooth even with 10 000 replications. The estimated root-MSEs have relatively large sampling variances, especially for large σ^2 . The

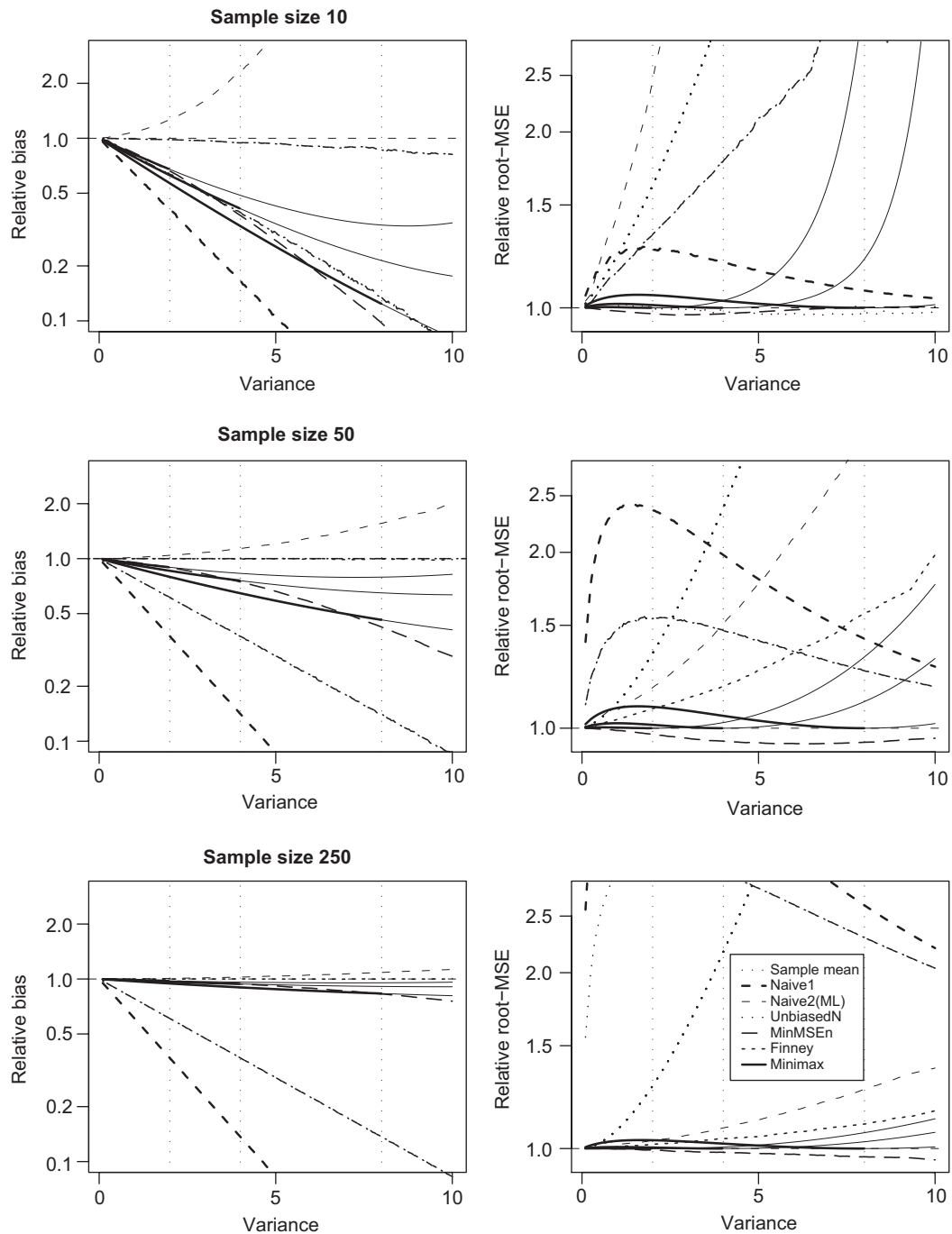


Fig. 2. The relative biases and relative root-MSEs of estimators of the expectation of the lognormal distribution, as functions of the variance σ^2 . The upper bounds σ_{mx}^2 are indicated by vertical dots.

uncertainty could be indicated in the diagram by pointwise confidence bands (error bars) but these would make the diagram too cluttered. The uncertainty is appreciable only for estimators in settings in which they are very inefficient. For large σ^2 and the naive estimators, the standard error $\sqrt{\text{var}\{\hat{\theta}(\hat{b}_a)\}}$ is more than ten times greater than the target $\theta(a)$. In these cases, the relative root-MSEs are estimated with standard errors that are in excess of 10% of their sizes.

The minimax estimators have negative biases, substantial for small sample sizes and large variances. On the log scale, the relative biases, drawn by thick solid lines in the intervals of plausible values of σ^2 , are indistinguishable from linear functions of σ^2 . Beyond σ_{mx}^2 , the negative bias gradually stops increasing and then decreases slowly, but the efficiency deteriorates, and does so rapidly for small samples and small values of σ_{mx}^2 .

With $\sigma_{\text{mx}}^2 = 8$, the minimax estimator is nearly as efficient as the ideal estimator even for $\sigma^2 = 10$, but for small values of σ^2 it is perceptibly inefficient. In contrast, when σ_{mx}^2 is set to 2 or 4, the loss of efficiency with respect to the ideal is only slight throughout $(0, \sigma_{\text{mx}}^2)$, but the loss is substantial for large values of σ^2 , especially with small samples. The minimax estimators have relatively small variances, and the biases are substantial contributors to the MSE for most values of σ^2 and $n \leq 50$. With increasing sample size, the root-MSE inflation when σ_{mx}^2 is understated becomes less severe.

The inefficiency of the ML estimator decreases with sample size, but at $n = 250$ it is still not competitive, except when σ^2 is very small. We conclude that $\hat{\theta}(\hat{b}_{0.5, \text{ms}}^*)$ is uniformly most efficient for $\theta(\frac{1}{2})$ among the estimators we explored for sample sizes in the range 15–250.

3.2. Estimators of the median

The median corresponds to the setting $a = 0$. The two naive estimators coincide for $a = 0$, so we use the label *Naive* for the common estimator. We also exclude estimator $\hat{\theta}(\hat{b}_{a, \text{ms}}^*)$ because for larger values of σ^2 (e.g., for $\sigma^2 > 5$ when $n = 50$) it attains very large values with nontrivial probabilities. The results for sample sizes $n = 10, 50$ and 250 are presented in Fig. 3. The biases of the estimators are much more moderate than for estimating the expectation $\theta(\frac{1}{2})$. We have proved the minimax property of the estimator $\hat{\theta}(b_{a, \text{mx}})$ only for $b > 0$. Even though $b_{0, \text{mx}} < 0$, the estimator $\hat{\theta}(b_{0, \text{mx}})$ has the minimax property. Together with the estimator $\hat{\theta}(\hat{b}_{0, \text{ms}}^*)$, it is about as efficient as the ideal estimator $\hat{\theta}(b_{0, \text{ms}}^*)$, and they are uniformly more efficient than the sample and naive estimators, as well as the estimators intended to be unbiased.

Some properties of $\hat{\theta}(\hat{b}_{0, \text{ms}}^*)$ and $\hat{\theta}(b_{0, \text{mx}})$ can be inferred from the approximation

$$D_0 = \exp \left\{ -\frac{3\sigma^2}{(k+2)n} \right\} \doteq 1 - \frac{3\sigma^2}{(k+2)n} \quad (14)$$

for $\sigma^2 \ll n^2$. The dependence of D_0 on σ^2 , or of $\hat{\theta}(\hat{b}_0)$ on $\hat{\sigma}^2$, is very weak for all but very small k and n . Of course, the sample median does not depend on $\hat{\sigma}^2$ at all, but the weak influence of $\hat{\sigma}^2$ on $\hat{\theta}(\hat{b}_0)$ is sufficient to make $\hat{\theta}(\hat{b}_0)$ an efficient estimator. For small $b\sigma^2$ (or $b\hat{\sigma}^2$), $\exp(b\hat{\sigma}^2) \doteq 1 + b\hat{\sigma}^2$ and $\text{var}(\exp(b\hat{\sigma}^2)) \doteq 2b^2\sigma^4/k$. That is why the root-MSEs of the estimators $\hat{\theta}(b)$ are approximately proportional to σ^2 . With the approximation in (14), $\hat{b}_0 \doteq -3/(2n)$, and therefore the log-relative bias is approximately proportional to σ^2 when $\sigma^2 \ll n$.

The sample median is uniformly the least efficient of the estimators we study, except for $n = 10$ and $\sigma^2 > 6$, because estimator $\hat{\theta}(\hat{b}_{0, \text{ms}}^*)$ has a breakdown at around $\sigma^2 = 6$, indicated by the vertical dashes. For greater sample sizes, this breakdown occurs for much greater values of σ^2 . For $n = 20$, $\hat{\theta}(\hat{b}_{0, \text{ms}}^*)$ is close to efficiency even for $\sigma^2 = 10$. With increasing sample size, the relative biases of all the estimators converge to zero. The minimax-like estimators are very forgiving, partly because $b_{0, \text{mx}}$ depends on σ_{mx}^2 very weakly; their inefficiency when $\sigma^2 > \sigma_{\text{mx}}^2$ is appreciable only for $n = 10$. We conclude by recommending the estimator $\hat{\theta}(\hat{b}_{0, \text{ms}}^*)$ when $n > 20$, although it is indistinguishable from $\hat{\theta}(b_{0, \text{mx}})$ based on a liberally chosen value of σ_{mx}^2 . For smaller sample sizes, $\hat{\theta}(b_{0, \text{mx}})$ does not have the drawback of a sudden MSE inflation.

3.3. Estimators of the mode

The mode of a continuous distribution does not have a natural sample or naive estimator. We could select a ‘window’ width w and define the estimator of the mode as the centre of the interval of width w (or the mean of such centres), which contains the largest number of observations. However, such an estimator is bound to be very inefficient.

The relative biases and root-MSEs of the estimators intended to be unbiased and efficient for $\theta(-1)$ and of the minimax-like estimators $\hat{\theta}(b_{-1, \text{mx}})$ are plotted in Fig. 4. The estimator $\hat{\theta}(\hat{b}_{-1, \text{ub}}^*)$ is severely biased and very inefficient, except when σ^2 is very small. The estimator $\hat{\theta}(\hat{b}_{-1, \text{ms}}^*)$ is biased somewhat less, and is much more efficient in ranges where it has a finite MSE. For $n = 10$, its MSE is infinite even for $\sigma^2 = 0.1$, and the corresponding breakdown point for $n = 50$ is at $\sigma^2 = 7.9$; the graph of its root-MSE is discontinued in the diagram at that value.

The biases and MSEs of $\hat{\theta}(\hat{b}_{-1, \text{ub}}^*)$ and $\hat{\theta}(\hat{b}_{-1, \text{ms}}^*)$ are extremely large; they are off the scale in all six panels of the diagram for most values of σ^2 , and are therefore not plotted at all. With a well informed setting of σ_{mx}^2 , the minimax-like estimator is quite efficient for $n \geq 50$. For $n = 50$, $\hat{\theta}(b_{-1, \text{mx}})$ with $\sigma_{\text{mx}}^2 = 8$ is only slightly less efficient than the ideal even for $\sigma^2 = 10$, and is more efficient than $\hat{\theta}(\hat{b}_{-1, \text{ms}}^*)$ for $\sigma^2 > 4.1$. It is relatively very inefficient only when σ^2 is very small. The estimators $\hat{\theta}(b_{-1, \text{mx}})$ with $\sigma_{\text{mx}}^2 = 2$ and 4 are quite efficient for very small values of σ^2 , but for large values of σ^2 they are very inefficient. The penalty for declaring a value of σ_{mx}^2 that is too large or too small is quite harsh.

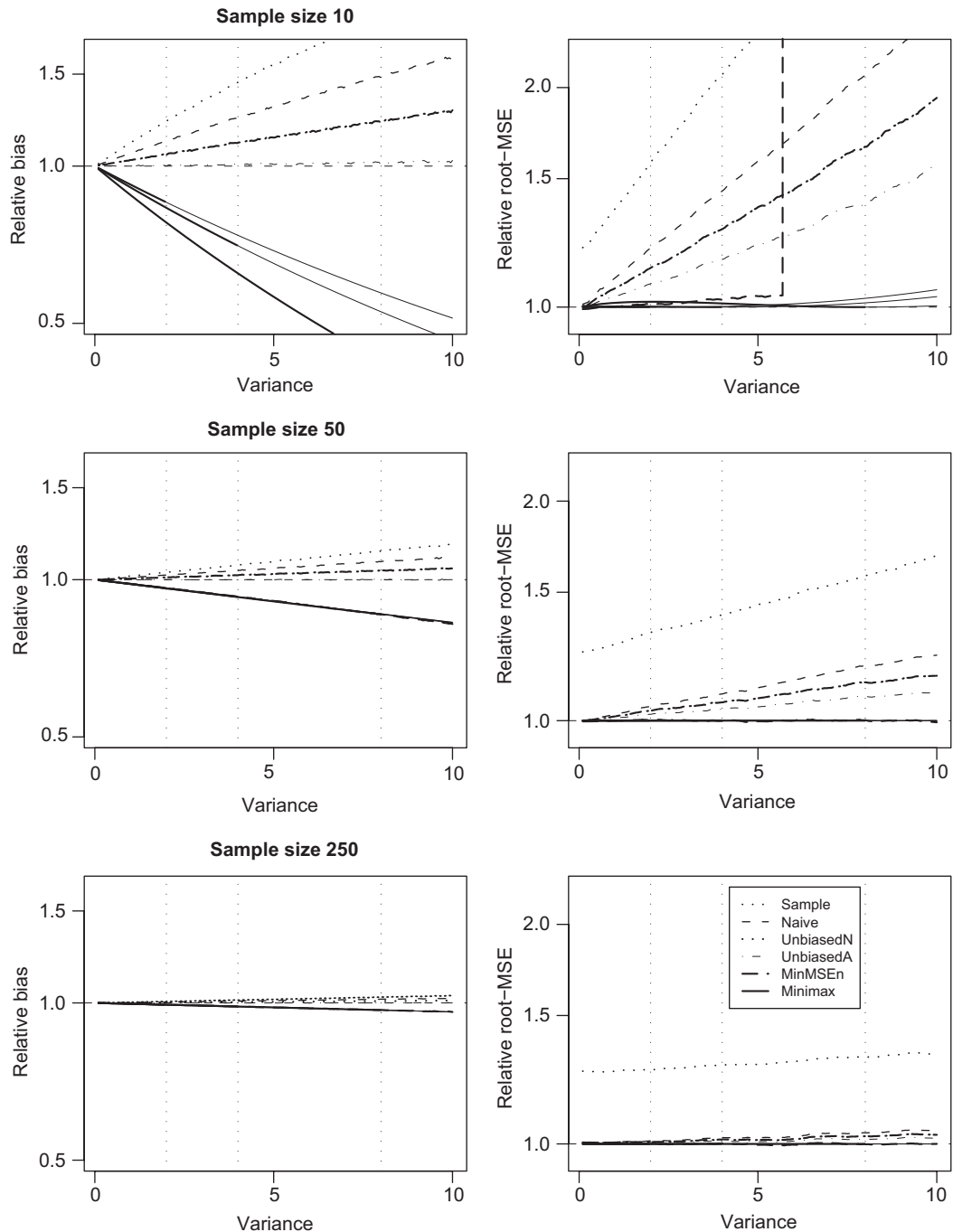


Fig. 3. The relative biases and relative root-MSEs of estimators of the median of the lognormal distribution, as functions of the variance σ^2 .

Some insight into the breakdown of the estimators $\hat{\theta}(\hat{b}_{-1,ms}^*)$ and $\hat{\theta}(\hat{b}_{-1,ms}^\dagger)$ can be gained directly from the expressions for the corresponding coefficients b . The singularity of the coefficient b_{-1} caused by the value $D_{-1} = \frac{1}{2}$ corresponds to $\sigma^2 = \log(2)(k+2)n/(2n+3)$, that is, for $n=50$ and $k=49$, to $\sigma^2 = 17.16$. In $\hat{b}_{-1,ms}^*$, we substitute $\hat{\sigma}^2$ for σ^2 in D_{-1} , so very large or very small values of \hat{b}_{-1} are obtained whenever values around 17 are plausible for $\hat{\sigma}^2$. For $n=50$ this occurs for $\sigma^2 > 7.9$. With the approximation to D_{-1} by the Taylor expansion, the denominator of $b_{-1,ms}^\dagger$ vanishes when $\sigma^2 = (k+2)n/(k(2n+3))$, which is close to 0.25 for all but very small k and n . Hence the breakdown of $\hat{\theta}(\hat{b}_{a,ms}^\dagger)$ for σ^2 around 0.25; it occurs also for other sample sizes. Breakdown

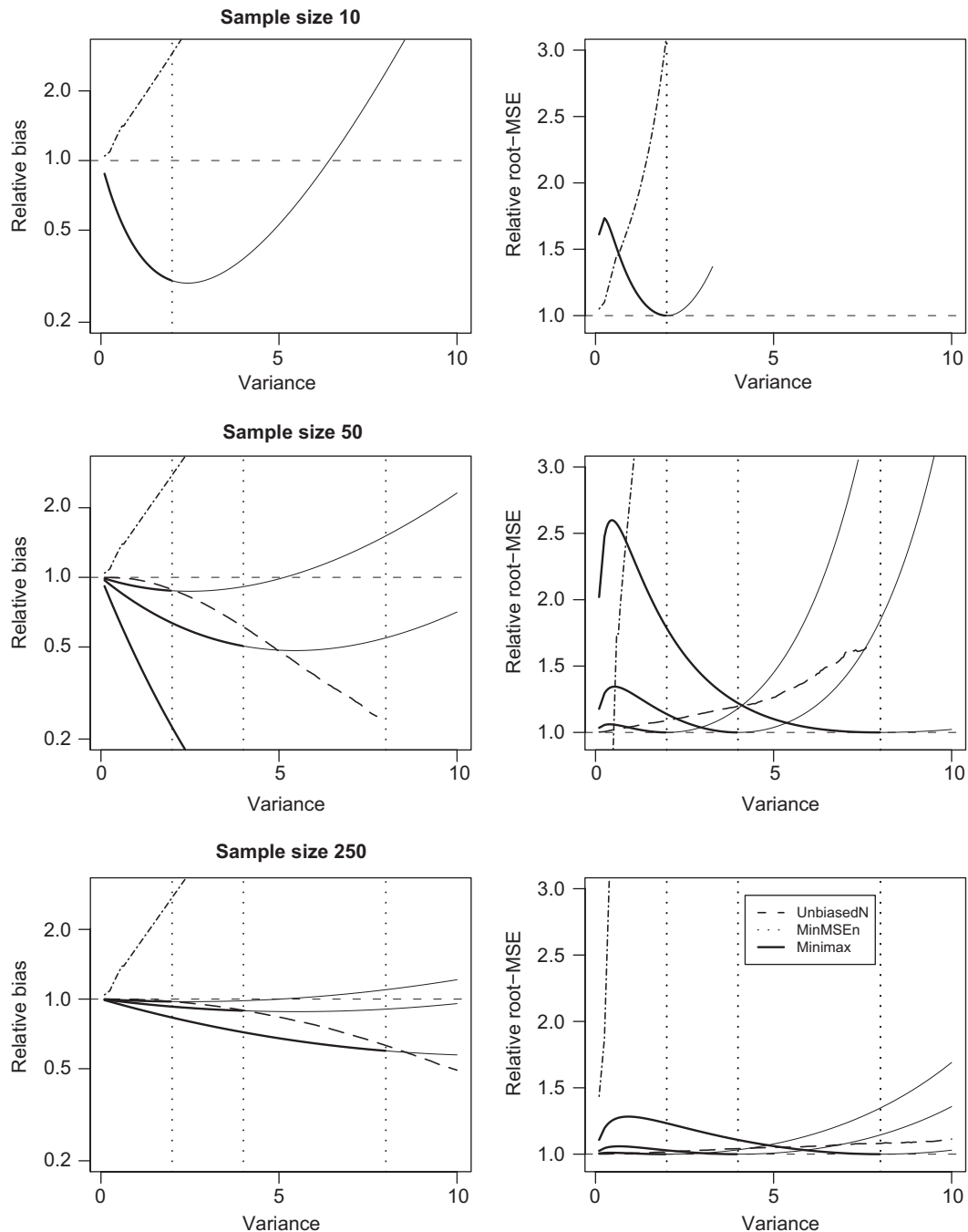


Fig. 4. The relative biases and relative root-MSEs of estimators of the mode of the lognormal distribution, as functions of the variance σ^2 .

occurs even for the minimax-like estimators for $\sigma_{\text{mx}}^2 = 4$ and 8. For sample size $n = 250$, only the estimator $\hat{\theta}(\hat{b}_{-1,\text{ms}}^*)$ performs well throughout the range of values of σ^2 .

We conclude by suggesting that estimation of the mode be based on at least $n = 50$ observations. Estimator $\hat{\theta}(\hat{b}_{-1,\text{ms}}^*)$ is nearly efficient for small values of σ^2 , and is a relatively safe choice otherwise. It breaks down for large values of σ^2 , but these breakdown values increase with σ^2 . If the values of σ^2 can be narrowed down, say, to an interval of length 2.0 or shorter, then the minimax-like estimator is quite efficient.

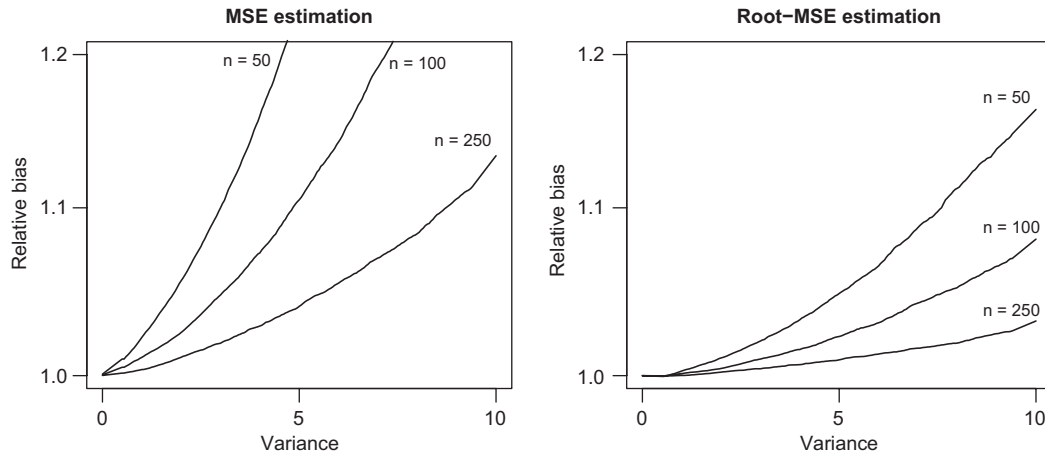


Fig. 5. The biases of the estimators of the MSE and root-MSE of the estimator $\hat{\theta}(\hat{b}_{0.5,ms}^*)$ of the expectation of the lognormal distribution, as functions of the variance σ^2 .

3.4. MSE estimation

In this section we discuss estimation of the MSE of $\hat{\theta}(\hat{b}_a)$, with $a = \frac{1}{2}$ and $\hat{b}_{0.5} = \hat{b}_{0.5,ms}^*$, for which the estimator is more efficient than the ideal. We estimate $MSE\{\theta(\hat{b}_{0.5,ms}^*); \theta(\frac{1}{2})\}$ by the MSE of the ideal ‘estimator’ evaluated at $\sigma^2 = \hat{\sigma}^2$, that is,

$$\hat{m}\left(\hat{b}_{0.5,ms}^*; \frac{1}{2}\right) = MSE\left\{\hat{\theta}(\hat{b}_{0.5,ms}^*); \theta\left(\frac{1}{2}\right)\right\} \Big|_{\sigma^2 = \hat{\sigma}^2}.$$

This is our only candidate for MSE estimation and we have no reference to which $MSE(\hat{m}; m)$ could be compared. We therefore adhere to the convention and study the bias of \hat{m} . Underestimation of MSE is usually regarded as a greater deficiency than overestimation by the same quantity; the former is interpreted as dishonest assessment, whereas the latter amounts to ‘underselling’ the quality of the estimator; see Longford (2007, Section A.20) for a discussion.

We have to consider estimation of MSE and root-MSE separately because these targets are related nonlinearly. For example, if an estimator \hat{m} is unbiased for m , then $\sqrt{\hat{m}}$ need not be unbiased for \sqrt{m} . The relative biases, defined as \hat{m}/m and $\sqrt{\hat{m}}/\sqrt{m}$, where the bar indicates averaging over the replications, are plotted for $n = 50, 100$ and 250 in Fig. 5. Both \hat{m} and $\sqrt{\hat{m}}$ overestimate their respective targets, the MSE and root-MSE, except for very small values of σ^2 . On the multiplicative scale, the extent of overestimation is smaller for root-MSE than for MSE, and is smaller for larger sample sizes. For $n = 10$ and 25 , the estimators are useful only for very small σ^2 ; for $n > 250$, the bias of the root-MSE estimator is very small even for $\sigma^2 = 10$.

The substantial bias of these estimators for small sample sizes should be judged in the context of large variance of the data as well as of the distribution of the MSE and rMSE estimators. For example, the relative biases of the MSE and root-MSE for $\sigma^2 = 5$ and $n = 10$ are 3.45 and 0.32, respectively, but these figures are associated with standard deviations (over the replications) of 35.6 and 1.11, respectively. We smoothed the empirical values of the biases only slightly, to retain an indication of the uncertainty about them that is present even in 10 000 replications.

4. Conclusions

Estimating the expectation, median and mode of the lognormal distribution are examples of failure of the maximum likelihood and of inapplicability of the asymptotic theory for sample sizes that for many other commonly encountered distributions would be sufficiently large. We derived estimators that are much more efficient than their naive alternatives, and explored how information about σ^2 can be incorporated in (minimax or minimax-like) estimation.

For estimating the expectation of the lognormal distribution from a random sample of size n , we propose the estimator $\exp(\hat{\mu} + \hat{b}\hat{\sigma}^2)$, where $\hat{\mu}$ and $\hat{\sigma}^2$ are the standard estimators for the respective mean and variance on the logarithmic (normal) scale, and

$$\hat{b} = \frac{n-1}{2\hat{\sigma}^2} \frac{D-1}{2D-1},$$

where

$$D = \exp\left\{\frac{n-3}{n(n+1)}\hat{\sigma}^2\right\}.$$

Although biased, this estimator is much more efficient than Finney's estimator, especially for large variances σ^2 . Finney's estimator is minimum-variance unbiased, a clearly formulated optimality property, whereas our estimator has no (universal) optimality properties. Our results indicate that unbiasedness and efficiency (small MSE) are conflicting inferential goals when estimating the location of a lognormal distribution. Insisting on unbiasedness when pursuing efficiency is an unaffordable luxury.

The counterintuitive result that $\hat{\theta}(\hat{b}_{0.5,ms}^*)$ is (slightly) more efficient than $\hat{\theta}(b_{0.5,ms}^*)$ for estimating the expectation $\exp(\mu + \frac{1}{2}\sigma^2)$ can be exploited for estimating the MSE of $\hat{\theta}(\hat{b}_{0.5,ms}^*)$ by substituting $\hat{\sigma}^2$ in the expression (9) for $\text{MSE}\{\hat{\theta}(b_{0.5,ms}^*); \theta(\frac{1}{2})\}$. The resulting MSE (and root-MSE) estimator has a positive bias which for a given σ^2 declines with sample size, and for a given sample size increases with σ^2 .

Our estimators rely on the functional form of the target, $\exp(\mu + a\sigma^2)$, as well as on independence of $\hat{\mu}$ and $\hat{\sigma}^2$, and so their robustness might be questioned. Robustness can be assessed by simulations, for instance, using the exponential of a distribution that differs slightly from the normal. For this purpose, we can use mixtures of normals with t distributions with few (or even only one) degree of freedom (and some noncentrality). We note that the contaminating (log- t) distributions do not have expectations and therefore neither do the mixtures for any number of degrees of freedom. In any sensitivity study, the naive estimators start with a considerable handicap which is unlikely to be overcome for moderate departures from lognormality.

Acknowledgements

Research for this manuscript was supported by the Grant SEJ2006-13537 from the Spanish Ministry of Science and Technology. Suggestions made by Omiros Papaspiliopoulos are acknowledged. An anonymous referee found several errors in the original submission and made very useful suggestions for how to improve the presentation.

References

- Aitchison, J., Brown, I.A.C., 1957. The Lognormal Distribution. Cambridge University Press, Cambridge, UK.
- Crow, E.L., Shimizu, K. (Eds.), 1988. Lognormal Distributions. Theory and Applications. M. Dekker, New York.
- Finney, D.J., 1941. On the distribution of a variate whose logarithm is normally distributed. J. Roy. Statist. Soc. 7 (Suppl.), 155–161.
- Limpert, E., Stahel, W.A., Abbt, M., 2001. Log-normal distributions across the sciences: keys and clues. Biosciences 51, 341–352.
- Longford, N.T., 2007. Studying Human Populations. An Advanced Course in Statistics. Springer, New York.
- Longford, N.T., Pittau, M.G., 2006. Stability of household income in European countries in the 1990s. Comput. Statist. Data Anal. 51, 1364–1383.
- Royston, P., 2001. The lognormal distribution as a model for survival time in cancer, with an emphasis on prognostic factors. Statist. Neerlandica 55, 89–104.
- Thöni, H., 1969. A table for estimating the mean of a lognormal distribution. J. Amer. Statist. Assoc. 64, 632–636.
- Toma, A., 2003. Robust estimators of the parameters of multivariate lognormal distribution. Comm. Statist. Theory Methods 32, 1405–1417.
- Zabel, J., 1999. Controlling for quality in house price indices. J. Real Estate Finance Econom. 13, 223–241.
- Zhou, X.-H., Gao, S., Hui, S.L., 1997. Methods for comparing the means of two independent log-normal samples. Biometrics 53, 1129–1135.